Article Watch: April 2019

Clive A. Slaughter

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Ave., Athens, GA 30606, USA. Tel: (706) 713-2216; Fax: (706) 713-2221; E-mail: cslaught@uga.edu, or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the Association.

AMINO ACID SEQUENCING

Swaminathan J, Boulgakov A A, Hernandez E T, Bardo A M, Bachman J L, Marotta J, Johnson A M, Anslyn E V, Marcotte E M. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nature Biotechnology* 36;2018:1076.

Swaminathan *et al.* explore an innovative approach to automated, ultrarapid, high sensitivity, proteomic analysis that combines aspects of traditional Edman chemistry, mass spectrometric database searching for peptide identification, and massively parallel, single-molecule DNA sequencing. The authors conduct massively parallel peptide sequencing by covalent labeling of peptides with fluorescent dyes that are amino-acid specific. The labeled peptides are immobilized on a glass coverslip; then their loss of fluorescence is monitored as amino acids are serially removed by successive cycles of Edman degradation. The process provides partial sequences from which the identity of proteins of origin can be inferred by database searching. Sequencing of peptides numbering thousands to millions could be performed on a single-molecule basis in this way. The setup is analogous to massively parallel DNA sequencing. The authors determine the positions of dye-labeled lysine and cysteine residues as well as fluorescently labeled phosphoserine within peptides. Simulations indicate that even with a 2-color code for lysine and cysteine, most proteins in mixtures of moderate complexity (~1000 proteins) can be uniquely identified from peptides generated by enzymatic cleave after glutamic acid or aspartic acid. The addition of dyes to label further amino acids such as tryptophan and aspartic or glutamic acid is expected to enable most proteins in the human proteome to be identified. The methodology in this paper is tested on zeptomole quantities of peptides. Because it operates inherently at single-molecule sensitivity, the procedure is capable of much better sensitivity than the attomole to femtomole (10^6–10^9 molecules) quantities required for analysis by mass spectrometry on a sensitive Orbitrap instrument. An enduring stumbling block for proteomics has been the enormous dynamic range required to analyze complex biologic mixtures (*e.g.*, 10^{12} for proteins in serum). The dynamic range requirement is much broader than that afforded by mass spectrometric detection (10^3–10^4). Dynamic range of single-molecule fluorescence sequencing is set by surface area of the flow cell, density of attached molecules, and imaging times. Contemporary Illumina DNA sequencers acquire hundreds of millions of reads per run, suggesting the possibility of an improvement in dynamic range with the envisaged methodology. Classic Edman degradation is limited in efficiency by repetitive yield and lagging because of partial cleavage. These problems are of lesser importance in the present system because single-molecule analysis renders population synchrony irrelevant, and adjustments for yield can be made in database searching. The pathway for further exploration of this promising technology is clarified by prior experience in protein chemistry, mass spectrometry, and DNA sequencing.

NUCLEIC ACID SEQUENCING

Bishara A, Moss E L, Kolmogorov M, Parada A E, Weng Z, Sidow A, Dekas A E, Batzoglou S, Bhatt A S. High-quality genome sequences of uncultured microbes by assembly of read clouds. *Nature Biotechnology* 36;2018:1067.

It has been estimated that <1% of microbial species are amenable to culturing. Knowledge of the remaining 99% depends on metagenomic sequencing of the complex communities in which the microbes occur. Metagenomic analysis involves the assembly of short-read sequences. The accurate assembly of complete genomes becomes very difficult when a community contains closely related species or species with high-copy repeat elements. Consequently, representative genomes for even some abundant taxa have

doi: 10.7171/jbt.19-3001-003

yet to be assembled. In the present paper, Bishara *et al.* adapt the "read cloud" assembly strategy introduced by 10x Genomics (Pleasanton, CA, USA) for use in metagenomics. They show that the methodology significantly out-performs existing metagenomic assembly procedures. The method utilizes droplet sequencing. High MW DNA fragments from a metagenomic sample are partitioned into droplets in an emulsion (~10 fragments/droplet) along with a barcoded primer for isothermal amplification within the droplet. The process yields short reads from the original long fragments that share a barcode called a "read cloud." The barcodes facilitate contig and, hence, genome assembly. The authors present a *de novo* assembler called Athena for use with metagenomic read clouds. During analysis of human fecal samples, the methodology produced larger numbers of complete genomes and higher contiguity in genome drafts than existing assembly methods. In an analysis of a complex marine sediment, it produced 9 nearly complete drafts, whereas simple short-read assembly produced none.

GLYCANS

Chen Y-H, Narimatsu Y, Clausen T M, Gomes C, Karlsson R, Steentoft C, Spliid C B, Gustavsson T, Salanti A, Persson A, Malmström A, Willén D, Ellervik U, Bennett E P, Mao Y, Clausen H, Yang Z. The GAGOme: a cell-based library of displayed glycosaminoglycans. *Nature Methods* 15;2018:881-888.

Qiu H, Shi S, Yue J, Xin M, Nairn A V, Lin L, Liu X, Li G, Archer-Hartmann S A, Dela Rosa M, Galizzi M, Wang S, Zhang F, Azadi P, van Kuppevelt T H, Cardoso W V, Kimata K, Ai X, Moremen K W, Esko J D, Linhardt R J, Wang L. A mutant-cell library for systematic analysis of heparan sulfate structure–function relationships. *Nature Methods* 15;2018:889-899.

Glycosaminoglycans (GAGs) are linear polysaccharides of complex structure found on cell surfaces and in the extracellular matrix. The cellular types (chondroitin sulfate, dermatan sulfate, and heparan sulfate) are involved in critical cell functions that include cell signaling, cell proliferation, tissue morphogenesis, and interactions with growth factors, cytokines, chemokines, and pathogens. The specific interactions between GAGs and protein ligands that serve these purposes are presumed to be mediated by binding sites that consist of relatively short tracts of variable modifications in specific arrangements, including *N*-sulfation, 6-*O*-sulfation, and 3-*O*-sulfation of glucosamine residues, 2-*O*-sulfation of uronic acid residues, and epimerization of glucuronic acid to iduronic acid. The glycan sequences and modifications are assembled by more than 40 glycosyltransferases, epimerases, sulfotranferases, and other enzymes. The salient structural motifs and the pathways by which they are assembled, however, are largely unknown, in part because of technical limitations in GAG structural analysis and lack of availability of GAG chains with defined structure.

Two groups now report the production by gene-editing techniques of panels of variant cells with stable changes in the enzymes that synthesize GAGs. The cells therefore synthesize GAGs with variant and novel structures that can be used in deciphering relationships between GAG structures and the cellular functions displayed by the variant cells. Chen *et al.*, working with Chinese hamster ovary cells, knock out specific combinations of genes using CRISPR–Cas9 and knock in other genes using zinc-finger nucleases. They demonstrate the use of their cell panel in studies of binding specificity for pathogens, growth factors, and antibodies. The cells produce GAGs with distinct structures available for the development of GAG microarrays. Qiu *et al.* are particularly interested in heparan sulfate. Because Chinese hamster ovary cells lack endogenous expression of many heparan sulfate–dependent signaling receptors and lack genes necessary for 3-*O*-sulfation, the authors choose mouse lung endothelial cells for construction of their variant panel. They employ Cre-*loxP* recombination, several existing mutant mouse strains, and CRISPR-Cas9 to produce variant forms of heparan sulfate. They dissect synthesis pathways, study antibody binding specificity, and provide information about the structural basis of fibroblast growth factor signaling. Qiu *et al.* resolve a longstanding question in the field in showing that heparan sulfate fine structure, not just overall sulfation level, determines biologic function. The work of these 2 groups provides resources that are expected to support ongoing advances in knowledge of selectivity in GAG-protein interactions.

METABOLOMICS

Gathungu R M, Larrea P, Sniatynski M J, Marur V R, Bowden J A, Koelmel J P, Starke-Reed P, Hubbard V S, Kristal B S. Optimization of electrospray ionization source parameters for lipidomics to reduce misannotation of in-source fragments as precursor ions. *Analytical Chemistry* 90;2018:13523-13532.

This report alerts investigators using electrospray mass spectrometry to misannotation as precursor ions of signals that actually represent the products of in-source

fragmentation of related endogenous metabolites. The work focuses on misannotation of lysophospholipids and phosphatidylcholines. Lysophospholipids can fragment to produce ions with the same mass as free fatty acids, lysophosphatidylcholines can fragment to produce ions with the same mass as free lysophosphatidylethanolamines, and phosphatidylcholines can fragment to produce ions with the same mass as phosphatidylethanolamines. Failure to interpret such signals correctly results in incorrect quantification or even assignment of a phenotype to the wrong lipid. The authors document instances of such confusion in the published literature and caution that shotgun lipidomic studies of complex samples are especially vulnerable. They suggest ways to recognize in-source fragmentation by inspection of chromatographic separation and ion intensities, and they suggest ion source conditions that reduce in-source fragmentation.

Binek A, Rojo D, Godzien J, Rupérez F J, Nuñez V, Jorge I, Ricote M, Vázquez J, Barbas C. Flow cytometry has a significant impact on the cellular metabolome. *Journal of Proteome Research* 18;2019: 169-181.

Metabolism of cells may undergo rapid and extensive changes under experimental conditions commonly used to separate cells. Binek *et al.* document such changes in mouse peritoneal macrophages during fluorescence-activated cell sorting. Most of the observed changes are associated with the flow cytometry process rather than with immunostaining. The authors observe changes in the abundance of glycerophospholipids, fatty-acid esters, amino acids and their derivatives, glycerolipids, and sphingolipids. They interpret the observations in terms of nutrient deprivation and the activation of signaling pathways in response to mechanical stress. The authors emphasize that these perturbations may occur in the absence of diminished cell viability. The results of the study are of interest both in the context of metabolomic investigations and in consideration of the cellular effects of *in vitro* or *ex vivo* experimental manipulations.

MACROMOLECULAR SYNTHESIS AND SYNTHETIC BIOLOGY

Cambray G, Guimaraes J C, Arkin A P. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nature Biotechnology* 36;2018:1005.

Cambray *et al.* undertake a massive study to disentangle the characteristics believed to affect translation efficiency in *Escherichia coli*. They construct 244,000 different variants of a 96-bp coding region fused to green fluorescent protein and measure reporter transcript abundance and decay, polysome profile, and cellular growth rate for each. The variants are designed to test the effects of global and local codon usage, mean hydropathy of the encoded amino acid sequences, and mRNA secondary structure at the 5′ end (around the translation start site), the middle, and the 3′ end of the coding region. The experimental design systematically assesses each of these variables in all possible combinations. Array synthesis and pooled sequencing are used in the compilation of a huge data set. Unexpectedly, transcript secondary structure, especially around the start codon, has the largest effect on translation efficiency. Codon usage also affects efficiency, but only when translation initiation is optimized and elongation becomes the limiting factor. Despite its comprehensive treatment of parameters known to affect protein production, the study explains only 53% of the total variance in protein yield, indicating that additional factors have yet to be identified.

MASS SPECTROMETRY

Dilillo M, de Graaf E L, Yadav A, Belov M E, McDonnell L A. Ultraviolet photodissociation of ESI- and MALDI-generated protein ions on a Q-Exactive mass spectrometer. *Journal of Proteome Research* 18;2019:557-564.

The singly charged molecular ions of peptides and proteins that predominate in matrix-assisted laser desorption ionization (MALDI) yield sequence-informative fragment ions by collisional dissociation or electron impact/transfer much less easily than do the multiply charged ions that arise in electrospray ionization. In the present paper, Dilillo *et al.* illustrate the utility of ulraviolet photon dissociation (UVPD) for generating informative fragments from singly charged ions. Using an Orbitrap Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a combined electrospray ionization and elevated pressure MALDI source (Spectroglyph, Kennewick, WA, USA), the authors implement UVPD and compare its utility in sequencing small proteins ionized by MALDI with in-source decay and higher energy collisional dissociation as alternative fragmentation methods. The authors conclude that UVPD provides the best sequence coverage for small proteins ionized by MALDI.

Liang S-Y, Patil A A, Han C-H, Chou S-W, Chang W, Soo P-C, Chang H-C, Peng W-P. Ionization of submicrometer-sized particles by laser-induced radiofrequency plasma for mass spectrometric analysis. *Analytical Chemistry* 90;2018:13236-13242.

Liang *et al.* devise a new method for desorbing and ionizing large particles such as viruses and bacterial cells for analysis by mass spectrometry. They employ a pulsed Nd:YAG laser beamed toward the backside of the sample substrate to generate acoustic waves for particle desorption. The laser pulse also triggers a radiofrequency plasma in a gas (argon, nitrogen, oxygen, methane, helium, xenon, or mixture thereof). Gas molecules ionize in the plasma, and the ions can participate in ion or molecule reactions with analyte particles in the gas phase to form multiply charged analyte ions. Average numbers of charges reach very high values as follows: 842 on freeze-dried *E. coli* EC11303, 1112 on *E. coli* strain W, and 971 on *Staphylococcus aureus*. A quadrupole ion trap operated at frequencies of several hundred hertz to a few kilohertz for charge detection mass spectrometry is used to measure mass-to-charge ratio values, which fall in the range 10^7–10^8, yielding average masses of 3.5×10^{10}, 6.0×10^{10}, and 5.6×10^{10} Da for the 3 bacterial types. Vaccinia virus with charge ~708 and m/z ~107 yields an average mass of $\sim 9.1 \times 10^9$ Da. The laser-induced radiofrequency plasma ion source is envisaged to extend the detectable mass range of native mass spectrometry.

Dominguez-Medina S, Fostner S, Defoort M, Sansa M, Stark A-K, Halim M A, Vernhes E, Gely M, Jourdan G, Alava T, Boulanger P, Masselon C, Hentz S. Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators. *Science* 362;2018:918-922.

Dominguez-Medina *et al.* adopt a different approach to measuring mass values in the range 10^6–10^{10} Da. They desorb particles from solution by surface acoustic wave nebulization at atmospheric pressure. The resulting mist of low kinetic energy droplets is passed through a heated metal capillary inlet, and the particles issuing forth are focused under their inertia in an aerodynamic lens without the need for electromagnetic fields. Their mass is measured using an array of nanomechanical resonators in a high vacuum. Individual particles landing on a resonator cause a shift in its resonance frequency that is a function of the particle's mass. Using this principle, the authors measure the mass of empty and DNA-filled bacteriophage T5 capsids up to 1.05×10^8 Da using <1 pmol of sample with an instrument resolution of >100. Either ionized or neutral particles can be analyzed with high sensitivity by this methodology. The authors anticipate that non-covalently associated components of particles may be less likely to dissociate than ions in mass spectrometric analysis, a prospective advantage in the study of large macromolecular assemblies.

FUNCTIONAL GENOMICS AND PROTEOMICS

Cao J, Cusanovich D A, Ramani V, Aghamirzaie D, Pliner H A, Hill A J, Daza R M, McFaline-Figueroa J L, Packer J S, Christiansen L, Steemers F J, Adey A C, Trapnell C, Shendure J. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361;2018:1380-1385.

Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, Hu G, Ni B, Zhao K. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* 562;2018:281-285.

Cellular responses involving changes in gene transcription are often driven by alterations in the accessibility of chromatin to binding by transcriptional regulators. Understanding of the way cells vary in chromatin architecture and the way accessibility may change with stimulus or damage is limited by the availability of techniques at the level of individual cells to study chromatin accessibility. Two groups contribute methods to address this need. Cao *et al.* determine jointly chromatin accessibility and transcriptional activity in single cells. They employ a "transposase-accessible chromatin" assay method (ATAC-Seq) to probe chromatin architecture. They apply the methodology to temporal changes in cells subjected to dexamethasone and to variations between mouse kidney cells, and they show that cis-regulatory sites may be identified by covariance of chromatin accessibility and transcription. Lai *et al.* jointly determine chromatin accessibility and nucleosome positioning in single cells. They employ a "micrococcal nuclease accessible chromatin" method (MN-seq) to deduce chromatin architecture. Studying NIH3T3 cells, mouse embryonic stem cells, and CD4+ T cells, they determine nucleosome positional variation and spacing at DNase-sensitive sites and enhancers.

Bycroft C, Freeman C, Petkova D, Band G, Elliott L T, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562;2018:203-209.

Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nature Genetics* 50;2018:1593-1599.

Bycroft *et al.* announce the release of data from the UK Biobank, a huge resource comprising both genomic and

clinical data that is available for discovery of genetic associations in common diseases. Study participants presently number 500,000. They donated samples of urine, saliva, and blood from which genotype information was acquired and biomarkers for known diseases were measured. Each participant also provided detailed information about lifestyle, medical history, and demographics. Some also underwent more detailed clinical testing. The participants were persons of European descent and 40–69 yr of age at the time of recruitment, which occurred between 2006 and 2010. They constitute a prospective cohort for assessment of development of adult diseases. Importantly, the UK Biobank is an open-access resource. Information about population structure and relatedness as well as genetic phasing and genotype imputation are also provided. The number of testable genetic variants is estimated to be 96 million. Canela-Xandri *et al.* use the UK Biobank to ascertain genetic associations for 660 binary traits and 118 nonbinary traits. The latter include continuous traits and traits with multiple ordered categories. Genetic data for over 9 million variants are included. Data for both significant and nonsignificant associations are provided in their study so that an unbiassed view of phenotype-genotype associations and correlations with environmental variables may be gleaned by investigators.

Martincorena I, Fowler J C, Wabik A, Lawson A R J, Abascal F, Hall M W J, Cagan A, Murai K, Mahbubani K, Stratton M R, Fitzgerald R C, Handford P A, Campbell P J, Saeb-Parsy K, Jones P H. Somatic mutant clones colonize the human esophagus with age. *Science* 362;2018:911-917.

Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, Shiozawa Y, Sato Y, Aoki K, Kim S K, Fujii Y, Yoshida K, Kataoka K, Nakagawa M M, Inoue Y, Hirano T, Shiraishi Y, Chiba K, Tanaka H, Sanada M, Nishikawa Y, Amanuma Y, Ohashi S, Aoyama I, Horimatsu T, Miyamoto S i, Tsunoda S, Sakai Y, Narahara M, Brown J B, Sato Y, Sawada G, Mimori K, Minamiguchi S, Haga H, Seno H, Miyano S, Makishima H, Muto M, Ogawa S. Age-related remodelling of esophageal epithelia by mutated cancer drivers. *Nature* 565;2019:312-317.

Cancer genetics has long distinguished between driver mutations, which confer the proliferative advantage that characterizes the malignant state, and passenger mutations, which may occur in normal or malignant cells and do not promote proliferation. The distinction is now blurred by the independent findings of 2 groups studying nonmalignant esophageal epithelia. Genome sequencing reveals that mutations in genes generally associated with malignancy accumulate even in nonmalignant esophageal epithelia. Such mutations appear in early childhood and increase in number with age. Clones bearing them undergo age-related expansion, indicating that they confer proliferative advantage. They occur in all persons of advanced age as an apparently normal result of aging. Nevertheless, the numbers of these pseudodriver mutations are much larger in esophageal squamous carcinoma cells than in nonmalignant cells, and their accumulation is accelerated in persons who smoke or consume alcohol heavily—behaviors that increase the risk of esophageal cancer. Interestingly, the prevalence of mutations in certain of the pseudodriver genes, notably *NOTCH1*, is much higher in nonmalignant than in malignant cells. These results raise questions about what genetic alterations drive the formation of cancers and how mutation in *NOTCH1* may contribute. High rates of somatic mutation are already well known in other tissues such as skin, and the accumulation of pseudodriver mutations with age now seems likely to be a feature of aging in many tissues.

PROTEOMICS

Kelstrup C D, Aizikov K, Batth T S, Kreutzman A, Grinfeld D, Lange O, Mourad D, Makarov A A, Olsen J V. Limits for resolving isobaric tandem mass tag reporter ions using phase-constrained spectrum deconvolution. *Journal of Proteome Research* 17; 2018:4008-4016.

This publication investigates data acquisition and processing conditions for multiplexed peptide quantification using tandem mass tags (TMTs). In 10-plex TMT reagents, the product reporter ions are closely spaced with a separation of 6.32 mDa. They therefore require high resolution to ensure accurate quantification. Because resolving power depends on acquisition time in Fourier transform mass analyzers such as Orbitraps, and a penalty is paid in proteome depth for increasing acquisition time, the authors determine the minimal resolution that is adequate. The required resolution also depends in part on the relative signal strengths of neighboring reporter ions. Using an Orbitrap Q Exactive HF-X mass spectrometer from Thermo Fisher Scientific (Bremen, Germany), the authors determine that a 64-ms transient acquisition time provides sufficient resolution to separate TMT reporter ions with absolute ratio changes of up to 16-fold (greater changes are seldom encountered). The advantage gained from the use of the recently described phased spectrum deconvolution method (ΦSDM) for signal processing of Fourier transform spectra is also assessed (see Grinfeld *et al.* Analytical Chemistry 89;2017:1202-1211). ΦSDM

produces interference-free mass spectra with resolution beyond the Fourier transform uncertainty limit. With only a small penalty in precision and accuracy, a 32-ms transient processed with ΦSDM deconvolution is shown to provide >50% more identifications, which are >99% quantifiable at a ratio limit of 16. The latter finding sets a new standard for multiplex TMT quantification studies.

CELL BIOLOGY AND TISSUE ENGINEERING

Biddy B A, Kong W, Kamimoto K, Guo C, Waye S E, Sun T, Morris S A. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* 564; 2018:219-224.

Direct cell lineage reprogramming, in which one differentiated cell type is converted to another without an intermediate pluripotent state, is typically an inefficient process that produces cells that incompletely recapitulate the target cell identity as well as cells of the intended type. Biddy *et al.* contribute new single-cell methodology to that already available for tracking the trajectories of cells undergoing direct lineage reprogramming with the intention of developing a better understanding of the sources of reprogramming inefficiency. In studies of the conversion of mouse embryonic fibroblasts to induced endodermal progenitors, they use a lentiviral vector for sequential delivery of 8-bp random barcodes to cells undergoing conversion. The transduced cells therefore express unique combinations of heritable tags that enable the construction of multilevel cell lineage trees based on single-cell RNA sequencing. The authors track over 1000 cells using this combinatorial indexing strategy and distinguish 2 trajectories, one leading to successful reprogramming and the other leading to a dead-end state. Which of the 2 paths is followed appears to be determined from the outset. This finding lends support to the suggestion that a privileged cell state exists in which reprogramming potential is predetermined. Successful reprograming is further associated with expression of a putative methyltransferase gene, *Mettl7a1*. Adding this gene to the reprogramming cocktail results in a 3-times increase in the number of successfully reprogrammed cells.

Moffitt J R, Bambah-Mukku D, Eichhorn S W, Vaughn E, Shekhar K, Perez J D, Rubinstein N D, Hao J, Regev A, Dulac C, Zhuang X. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362;2018: eaau5324.

Single-cell RNA-seq is incomparably powerful for distinguishing cell types and cataloging their abundance in tissues. Because the technique requires dissociation of cells, however, it fails to provide information about the spatial relationships between the cells. Moffitt *et al.* here combine single-cell RNA-seq with a tissue-imaging approach, FISH, to characterize the spatial relationships between cell types at the single-cell level. Their interest lies in understanding the function of the preoptic region of the mouse hypothalamus, which controls social behaviors such as parenting, mating, and aggression as well as homeostatic functions such as thermoregulation, thirst, and sleep. They use the information acquired by single-cell RNA-seq for 31,000 cells to choose probes for FISH best suited to distinguish the cell types identified. Using these probes, they then image 1.1 million cells using multiplexed error robust FISH (Moffitt *et al.* Proc Natl Acad Sci USA 113;2016:11046-11051; Moffitt *et al.* Proc Natl Acad Sci USA 113;2016:14456-14461) to map spatial organization of the cell types identified by RNA-seq. The process reveals ~70 distinct neuronal populations distinguished by their transcriptional diversity and spatial clustering. The sensitivity of multiplexed error robust FISH further provides an opportunity to study the activation of immediate early genes in response to specific behaviors. The authors document involvement of different neuronal populations in parenting, aggression, and mating. The methodology is anticipated to contribute extensively to studies of the neural basis of normal and abnormal behaviors.

IMAGING

Sieben C, Banterle N, Douglass K M, Gönczy P, Manley S. Multicolor single-particle reconstruction of protein complexes. *Nature Methods* 15;2018:777-780.

Heydarian H, Schueder F, Strauss M T, van Werkhoven B, Fazel M, Lidke K A, Jungmann R, Stallinga S, Rieger B. Template-free 2D particle fusion in localization microscopy. *Nature Methods* 15;2018:781-784.

The 3-dimensional structures of macromolecular complexes may be ascertained using multiple fluorescence images of single particles. To make use of multiple images for reconstructing structure, methods for aligning the individual images and then averaging them must be deployed to deal with issues of random particle orientation, incomplete fluorophore labeling, and low signal strength. Two groups now contribute to the computational methods available for such analyses. Sieben *et al.* are studying the architecture of the human centriole, a particle involved in the formation of cilia, flagella, and centrosomes, composed

of >100 proteins organized into various substructures. The authors acquire 2-dimensional single-molecule localization microscopy data sets. They make alignments based on one of the polypeptide components, which they treat as a reference protein throughout, and they localize the various other components relative to that reference. Heydarian *et al.* developed an all-to-all particle registration routine. This method avoids bias introduced with the introduction of a structural template for alignment. It also minimizes effects of under-labeling and misregistration. The authors reconstruct an 8-fold symmetric ring structure for the nuclear pore complex based on stochastic optical reconstruction microscopy images of the integral membrane protein component gp120. The software developed by both groups is made available to assist other investigators.

Guo Y, Li D, Zhang S, Yang Y, Liu J-J, Wang X, Liu C, Milkie D E, Moore R P, Tulu U S, Kiehart D P, Hu J, Lippincott-Schwartz J, Betzig E, Li D. Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales. *Cell* 175;2018:1430-1442.e1417.

New imaging technologies for studying the dynamic reorganization of organelles and cytoskeleton have revealed interactions of contact between subcellular structures that control fundamental cellular functions such as mitochondrial DNA synthesis, endosomal fission, and late endosome transport by molecular motors. However, limitations in the resolution, speed, and *z*-depth of imaging have constrained development of knowledge in this arena. Guo *et al.* have now developed imaging methodology that affords substantial improvement in capability. It enables dynamic events to be followed within the portion of the endoplasmic reticulum (ER) close to the basal plasma membrane at a 97-nm resolution and 266 frames per s over thousands of time points. This is accomplished with an improvement to superresolution structured illumination microscopy, in which the illumination is launched just inside the critical angle for total internal reflection fluorescence (grazing incidence), creating an illumination field nearly parallel to the substrate to which cells are attached, with thickness comparable to the objective's depth of focus. This innovation maximizes the *z*-depth of imaging while eliminating out-of-focus background. The system enables live cell imaging with high spatial and temporal resolution but low photobleaching and phototoxicity. The enhanced capabilities enable the authors to discover new mechanisms by which tubular ER is generated. They also deduce that ER is reshaped by hitchhiking on molecular motors when the motors move organelles to which the ER becomes attached. They further observe that contacts between ER and mitochondria can induce mitochondrial fusion as well as fission, and that collisions between ER and late endosomes or lysosomes can cause breakage of ER tubules. The authors anticipate further improvements in capability of the system by which to probe dynamic interactions within cells.

POLICY

Erlich Y, Shor T, Pe'er I, Carmi S. Identity inference of genomic data using long-range familial searches. *Science* 362;2018:690-694.

Readers will have noted the recent use of genotype data from a crime scene sample to trace a suspect in the Golden State Killer case through an online familial search provided by GEDmatch (Lake Worth, FL, USA) as a direct-to-consumer service. Such long-range family searches (*i.e.*, searches designed to identify distant relatives) are not subject to legal regulation. Subsequent activity by law enforcement agencies suggests that the use of such online services will become standard forensic practice. Erlich *et al.* here estimate the power of long-range family searches to identify targeted individuals using databases presently available. For persons of European descent, they estimate that 60% of searches presently result in identification of at least 1 third cousin or a closer relative. As the consumer databases increase in size, the probability of matching at the third-cousin level increases, as does the probability of matching at the second-cousin level. With the addition of demographic data such as age, in the foreseeable future, specific individuals will be identifiable from lists of matches returned by searches. This prospect presents incipient issues of genetic confidentiality associated with long-range family searching. For example, research subjects are likely to become identifiable from their genetic data. Such issues will be of interest and concern to many readers of *JBT*.